

# Patient Empowerment through Summarization of Discussion Threads on Treatments in a Patient Self-Help Forum

Sourabh Dandage<sup>\*1</sup>, Johannes Huber<sup>\*1</sup>, Atin Janki<sup>1</sup>, Uli Niemann<sup>1</sup>, Ruediger Pryss<sup>2</sup>, Manfred Reichert<sup>2</sup>, Steve Harrison<sup>3</sup>, Markku Vessala<sup>3</sup>, Winfried Schlee<sup>4</sup>, Thomas Probst<sup>5</sup> and Myra Spiliopoulou<sup>1</sup>

<sup>1</sup> Otto-von-Guericke Univ. Magdeburg, Germany – first authors with \* have equal contribution

<sup>2</sup> University of Ulm, Germany

<sup>3</sup> TinnitusHub, UK

<sup>4</sup> University Hospital Regensburg, Germany

<sup>5</sup> Donau Univ. Krems, Austria

**Abstract**—Self-help patient fora are widely used for information acquisition and exchange of experiences, e.g., on the effects of medical treatments for a disease. However, a new patient may have difficulties in getting a fast overview of the information inside a large forum. We propose TinnitusTreatmentMonitor, a prototype tool for the summarization and sentiment characterization of postings on medical treatments. We report on applying TinnitusTreatmentMonitor on the platform TinnitusTalk<sup>1</sup>, a self-help platform for tinnitus patients.

**Keywords**— self-help patient fora, opinions on treatments, discussion threads, sentiment analysis, medical mining

target extraction (e.g., [6, 7]). Our approach is partially inspired by [7], which identifies drugs discussed in opinionated postings and also detects subjectivity. Apart from these methods, TinnitusTreatmentMonitor also investigates polarity evolution for the studied treatments.

Our contribution is a proof-of-concept framework that gives an overview of discussions on medical treatments. It encompasses components for the recognition of treatments and of the dominant polarity associated with each treatment at each timepoint. We present the framework and used materials in the next sections. Our results and discussion are presented thereafter. We close the paper with a summary and outlook.

## I. INTRODUCTION AND RELATED WORK

Self-help internet fora allow patients to share experiences on their disease. However, a forum may contain a huge number of discussion postings and new users may have difficulties in acquiring a fast overview of the discussed contents. We propose TinnitusTreatmentMonitor, a framework that gives users a fast overview of discussions on tinnitus treatments.

Tinnitus is defined as the condition of hearing sounds without external stimulus. According to [1], tinnitus prevalence is 10-15%, while 1-2% of the patients experience a deterioration of quality of life. Insights on potential therapies are intensively discussed in social platforms like TinnitusTalk. This platform was established in March 2011 and supports discussions of treatments, exchange of experiences and support. Its two subfora on treatments contained (in July 2017) more than 35,000 postings by approximately 1100 authors [2].

The analysis of discussions in online patient platforms is intermittently done manually. For example, [3] focuses on information correctness. Most frequently, machine learning is used though, as in [4, 5, 6, 7]. Relevant to our approach are the tasks of sentiment analysis (e.g., [5]), and opinion

## II. MATERIALS

For our analysis, we used the 9 TinnitusTalk subfora listed in Table 1. Thereby, the Postings column counts postings referring to treatments, the Mentions column counts sentences mentioning treatments.

Table 1: Materials from 9 subfora of TinnitusTalk (collected in July 2017)

#	Subforum	Postings	Mentions	Authors
1	Alternative Treatments	15990	8188	710
2	Collaboration Space	126	87	21
3	Introduce Yourself	27595	11365	1854
4	Research News	24722	6776	640
5	Success Stories	8110	2199	522
6	Support (Tinnitus)	121320	28893	2062
7	Support (Pulsatile T.)	2889	457	134
8	Support (Hyperacusis)	6965	2328	254
9	Treatments	19657	15523	1108
Total		227374	75816	3950

## III. TINNITUSTREATMENTMONITOR

In Fig. 1, we show the workflow of TinnitusTreatmentMonitor. Tasks are in dark blue and outputs in light blue.

<sup>1</sup>TinnitusTalk.com, operated by TinnitusHub.com

The back-end consists of the components for the tasks of data collection, identification of sentences mentioning treatments, sentence labeling, aggregation and scoring. They were implemented in Python, using the external libraries *Scrapy* for crawling, *NLTK*[8] for text processing *scikit-learn*[9] for classification and *Pandas*[10] for data aggregation. The front-end task of visualization was implemented in Javascript, using the *Aurelia* framework: it acquired inputs from an API built on top of the *Tornado* server and used *Bokeh* [11] for graph rendering. We describe all tasks hereafter.

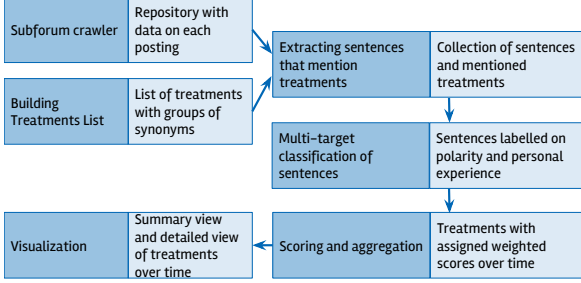


Fig. 1: The workflow and tasks of TinnitusTreatmentMonitor

#### A. Subforum crawler

In each subforum, the crawler extracts the ID, timestamp and author ID of each posting, the thread containing the posting and the number of users who clicked the button “agree” for it. It also stores the text (no images) after removing double newlines and quotations of earlier postings.

#### B. Building a list of treatments

This component takes a handcrafted list of treatment names as input: we extracted them from the titles of posting threads in the subfora “Treatments” and “Alternative Treatments and Research”. A treatment is mentioned with multiple names: we traced those synonyms and grouped them together. During the task of labeling a sample of sentences manually (see Section III.D), we identified further names and added them to the list, completing with 149 names for 48 different treatments in total.

#### C. Extracting sentences that mention treatments

This component first splits each posting into sentences, using the “Punkt” sentence tokenizer of [12]. If a sentence  $s$  mentions a treatment, it is stored together with the treatment(s) it refers to. The two subsequent sentences are also inspected; if they do not mention a different treatment, they are stored with  $s$  and jointly considered during classification.

#### D. Multi-target classification of sentences

For this task, we train and apply a *multi-target classifier* on two target variables, namely “polarity” and “personal experience”. To this purpose, we model the sentences as vectors of derived features, which we compute using natural language processing tools and lexical resources, including [13, 14, 15, 16]. For the target variable “polarity”, we define the labels “positive”, “negative” and “neutral”. The target variable “personal experience” specifies whether the author of the posting discusses the treatment on the basis of the own personal experience, hence we define that this target variable has the values YES and NO.

We created a random sample of 600 sentences and labeled them, splitting into a training set of 400 sentences and using the rest for testing. We use a multi-target random forest classification algorithm, the Python *scikit-learn* [9] implementation of [17]. Once the multi-target classifier is learned, we apply it to all sentences extracted in the previous task.

#### E. Scoring and aggregation

Building upon the labels assigned to each sentence, this component computes scores and aggregates them for each treatment and timeframe. In particular, a weighted score is computed for each sentence, by mapping the polarity labels “positive”, “negative” and “neutral” to the *scores* 1, -1 and 0 respectively, and the personal experience labels to the *weights* 1.0 (label YES) and 0.2 (label NO). Then, for each posting  $x$  and treatment  $y$ , where  $S(x, y)$  is the set of sentences in  $x$  that refer to  $y$ , we compute:

$$pScore(x, y) = \frac{\sum_{s \in S(x, y)} score(s) \cdot weight(s)}{\sum_{s \in S(x, y)} weight(s)} \quad (1)$$

Each posting acquires a weight depending on the maximum sentence weight in the posting and the number of users clicking the “agree”-button for it:

$$pWeight(x, y) = \max_{s \in S(x, y)} weight(s) \cdot (1 + 0.5 \cdot agrees(x)) \quad (2)$$

For each time period and treatment, the weighted average of the associated posting scores is calculated, stored and presented as “treatment score” for this period. Currently, we support two time granularities, month and year.

#### F. Visualization

The last component of TinnitusTreatmentMonitor is an interactive web application that shows how each treatment is

mentioned and perceived in the forum. It consists of a “summary view” over all treatments and a “detailed view” for each treatment chosen by the user.

#### IV. RESULTS

We run TinnitusTreatmentMonitor on TinnitusTalk. Of the postings recorded till July 2017, 41,193 (written by 3,950 users) mentioned treatments. These mentions were in 75,816 sentences, 12,979 written in the last year. We identified a negative tendency in the users’ opinions: only 9 of the 48 treatments had treatment scores with a positive average.

On the testing subsample of our manually annotated sample of statements, our multi-target classifier achieved an accuracy of 60% for the polarity target ,and 68% for the personal experience target. We compared to a baseline that assigns to each statement the label of the majority class: its accuracy was 57% for polarity ,and 52% for personal experience, hence our model improved the baseline.

Fig. 2 depicts the “summary view”, which is also the start page of the front end of TinnitusTreatmentMonitor. It contains one row per treatment, consisting of three tiles: the names associated with the treatment (leftmost tile), the number of mentions in the last period (middle tile) and the treatment score (rightmost tile). This view also allows that the user filters out treatments or sorts them.

Search...	Frequency	
<b>Antidepressants</b> Amitriptyline, Celexa, Citalopram, Cymbalta, Effexor, Fluoxetine, Lexapro, Mirtazapine, Nortriptyline, Paxil, Pristiq, Prozac, Remeron, SSRI, SSRIs, Sertraline, Venlafaxine, Wellbutrin, Zoloft, anti-depressant, anti-depressants, antidepressant	<b>1804</b> mentions in the last 12 months	<b>-0.11</b> treatment score
<b>Masking</b> masker, maskers	<b>1729</b> mentions in the last 12 months	<b>-0.01</b> treatment score
<b>Steroids</b> Dexamethasone, Lidocaine, Medrol, Prednazone, Prednisolone, Prednisone	<b>1686</b> mentions in the last 12 months	<b>-0.07</b> treatment score
<b>Tinnitus Retraining Therapy</b> TRT, retraining	<b>1543</b> mentions in the last 12 months	<b>-0.03</b> treatment score
<b>Low-level Laser Therapy</b>	<b>1270</b>	<b>0.02</b>

Fig. 2: Summary View

By clicking on the tiles at the right side of the “summary view” for some treatment, the human expert comes to the “detailed view” of a treatment. Fig. 3 depicts one treatment. Above the graph, we see its names. To the right of the graph, the tiles from top to bottom display the absolute number of mentions, the percentage of mentions and the treatment score. The arrow indicates the respective trend over the last year.

The main part of the Fig.3 offers three interactive graphs: each can be chosen by clicking at the corresponding tile. The graph chosen on Fig. 3 shows the absolute number of

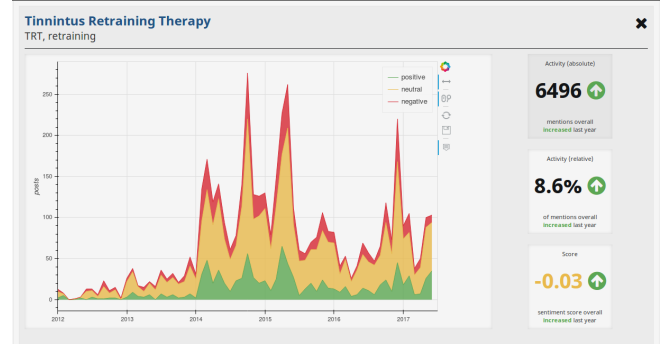


Fig. 3: Detailed View

mentions over time, with positive, neutral or negative sentiment indicated by green, yellow, resp. red color. The “detailed view” also includes a choice of threads and sentences per treatment (not shown in Fig. 3). In particular, for each treatment, the posting threads are ranked on the number of the treatment’s mentions in them and the top-5 threads are shown. Within each of these threads, the sentences are ranked on the target variable of personal experience (YES preferred over NO), on the number of agrees for the posting and the timestamp of the posting. The top-5 sentences are presented, with links to the original postings, so that the user can follow the links as entry points into the corresponding subfora.

#### V. DISCUSSION

TinnitusTreatmentMonitor aims at providing users with a comprehensive treatment overview and how these treatments are perceived over time. The presented back-end components contribute to this goal by identifying the postings mentioning treatments, classifying them and eventually visualizing them. The front-end visualization is based on the back-end and complements TinnitusTreatmentMonitor . The summary view of the visualization component (cf. Fig. 2) assists users in obtaining a first impression of all treatments. Using the leftmost column, users can figure out which names are used for the treatment. In the middle column, users see whether a treatment is subject of vivid discussions (large number of recent mentions) or it is stalled (small number). From the rightmost column, users can conclude whether the treatment is perceived positively or negatively. Hence, for new users, the summary view provides a first and compact impression of all discussed treatments. For users with a specific treatment in mind, the summary view allows them to compare its statistics with those of other treatments.

The detailed view builds upon the first overview to assist a user in understanding how treatments are perceived. By jux-

taping the area under each color in the graph of Fig. 3, the user can decide easily on whether the discussion on the specific treatment is mostly neutral (when the yellow area is predominant) or mostly sentimental (when the red or green areas are predominant). Hence, the user gets a first impression on the opinion of other patients on the treatment over time, without needing to read the postings. Hence, our proof-of-concept prototype can assist new users in acquiring insights on the discussion intensity and sentiment for all treatment categories and for each treatment separately. In a next step users will get the opportunity to evaluate the prototype in our lab and then in a sandbox of TinnitusTalk. Note that our approach revealed also drawbacks that must be further addressed. First, treatments are collected manually and, hence, we plan to use methods for Opinion Target Extraction [18, 19]. Next, the label assignments done by the classifier after training are not verified by a human expert and therefore we intend to address this task with active learning methods (see [20] on a stream of postings). Finally, the approach does not take the user's interests into account. A first step constitutes a personalized search by acquiring, e.g., keywords from the user or identifying treatments where only these keywords show up.

## VI. SUMMARY AND OUTLOOK

We proposed a method for the compact representation of postings in a social platform for patient self-help. Thereby, we focused on treatments and developed mechanisms, which assess the polarity of postings referring to a treatment, visualizing popularity and polarity of each treatment across the time axis, and also providing aggregated information over all treatment categories. Our next steps include the automation of treatment recognition from texts, and semi-supervised or active learning for the acquisition of human-verified polarity labels. Regarding user acceptance, we anticipate a first study with users interacting with our environment for different search tasks. The ultimate goal of our approach is to assist users in finding information. Hence, we plan to extend the approach with a keyword-based querying mechanism allowing users to learn about popularity and polarity trends for specific treatments. The information returned can be used in different ways. Of particular interest is offering it to a patient via a mobile self-help app.

## COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest and no conflict with ethical standards. The social platform is public domain.

## ACKNOWLEDGEMENTS

Partly, the work done by U. Niemann and M. Spiliopoulou was within the German Research Foundation project OSCAR “Opinion Stream Classification with Ensembles and Active Learners”: U. Niemann is partially funded by OSCAR, whereas M. Spiliopoulou is project investigator.

## REFERENCES

1. Baguley D, McFerran D. Tinnitus *The Lancet*. 2013;382:1600–1607.
2. Probst Thomas et al.. Outpatient Tinnitus Clinic, Self-Help Web Platform, or Mobile Application to Recruit Tinnitus Study Samples? *Frontiers in aging neuroscience*. 2017;9:113.
3. Türp Jens, Ohla Harald. Temporomandibular joint pain: Analyzing discussions in online forums *Zeitschrift für Krianiomandibuläre Funktion*. 2012;4:227–244.
4. Liu Xiao, Chen Hsinchun. *AZDrugMiner: An Information Extraction System for Mining Patient-Reported Adverse Drug Events in Online Patient Forums*:134–150. Berlin, Heidelberg: Springer Berlin Heidelberg 2013.
5. Korkontzelos I. et al.. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts *Journal of Biomedical Informatics*. 2016;62:148 - 158.
6. Lorraine G. et al.. Sentiment Lexicons for Health-related Opinion Mining in *Proc of the 2Nd ACM SIGHIT Int'l Health Informatics Symp*:219–226ACM 2012.
7. Asghar Dr. Muhammad et al.. Health miner: opinion extraction from user generated health reviews 2013;5:279-284.
8. Bird Steven et al.. *Natural Language Processing with Python*. O'Reilly1st ed. 2009.
9. Pedregosa F. et al.. Scikit-learn: Machine Learning in Python *Journal of Machine Learning Research*. 2011;12:2825–2830.
10. McKinney Wes. *Python for data analysis*. O'Reilly1st ed. 2012.
11. Bokeh Development Team . *Bokeh: Python library for interactive visualization* 2014.
12. T Kiss, J Strunk. Unsupervised Multilingual Sentence Boundary Detection *Computational Linguistics*. 2006;32:485-525.
13. Hu Minqing, Liu Bing. Mining and summarizing customer reviews in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*:168–177ACM 2004.
14. Motoda Hiroshi et al.. Advanced Data Mining and Applications *Lecture Notes in Artificial Intelligence*. 2013;1:XXII, 588.
15. Hutto Clayton J, Gilbert Eric. Vader: A parsimonious rule-based model for sentiment analysis of social media text in *8th Int. AAAI Conf. on Weblogs and Social Media* 2014.
16. Jiang Keyuan et al.. Construction of a Personal Experience Tweet Corpus for Health Surveillance *ACL 2016*. 2016:128.
17. Louppe Gilles. Accelerating Random Forests in Scikit-Learn 2014.
18. Deng Lingjia, Wiebe Janyce. Joint Prediction for Entity/Event-Level Sentiment Analysis using Probabilistic Soft Logic Models in *2015 Conf on Empirical Methods in Natural Language Processing* Association for Computational Linguistics 2015.
19. J Niklas, I Gurevych. Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields in *Conf on Empirical Methods in Natural Language Processing*:1035-1045 Association for Computational Linguistics 2010.
20. Zimmermann Max et al.. Incremental Active Opinion Learning Over a Stream of Opinionated Documents in *WS on Issues of Sentiment Discovery and Opinion Mining at KDD* 2015.